

# Stichprobenumfang und Fragebogenlänge in Webevaluationen

Gerrit Hirschfeld<sup>1</sup> (gerrit.hirschfeld@fh-bielefeld.de), Meinald T. Thielsch<sup>2</sup> (Thielsch@uni-muenster.de)

1. Fachhochschule Bielefeld, Bielefeld, Deutschland
2. WWU Münster, Münster, Deutschland

## Einleitung

Oftmals hört man in der Praxis die von Jacob Nielsen (1993) geprägte Daumenregel fünf Testpersonen seien ausreichend für Webevaluationsstudien. Derartige Daumenregeln sind hilfreich für explorativ-qualitative Usability-Tests, aber hinsichtlich einer fragebogenbasierten Website-Evaluation lässt sich der notwendige Stichprobenumfang statistisch exakt bestimmen.

## Minimale Stichprobenumfänge

Im Falle von Evaluationsstudien von Webseiten ist es das Ziel, möglichst genau ein bestimmtes Merkmal (etwa die Bewertung des Inhalts) zu schätzen. Diese Genauigkeit kann man am einfachsten anhand der Breite des Konfidenzintervalls beschreiben.

**Konfidenzintervalle** geben einen Bereich um den empirisch gefundenen Mittelwert an, innerhalb dessen in 95% der Fälle der wahre Wert liegt. Konfidenzintervalle werden mit zunehmendem Stichprobenumfang kürzer und somit genauer, liegen also näher um den wahren Wert. Neben der gewünschten Genauigkeit ist auch die Variabilität des Merkmals relevant. Je stärker dieses Merkmal schwankt, desto mehr Probanden müssen befragt werden (Kauermann and Küchenhoff 2011)

Die Schwankung der Messwerte wiederum setzt sich zusammen aus der **Reliabilität** und der tatsächlichen Standardabweichung (Trafimow 2014). Je höher die Reliabilität des Messinstruments ist, desto geringer ist die Schwankung der Messwerte und umgekehrt.

Zuletzt kann man einen Zusammenhang herstellen zwischen der Itemanzahl und der Reliabilität des Messinstruments. So kann man anhand der Spearman-Brown-Vorhersageformel (Nunnally, Bernstein, and Berge 1967) schätzen, wie sich die Reliabilität eines Fragebogens verändert, wenn man die Länge variiert.

Diese Zusammenhänge sollen am Beispiel des Web-CLIC verdeutlicht werden.

## Webinhalt - Web-CLIC

Der Inhalt ist das zentrale Element einer Website – und wird von NutzerInnen als das wichtigste Kriterium für die Beurteilung genannt (Thielsch, Blotenberg, and Jaron 2014). Der Web-CLIC ist ein Fragebogen, mit dem NutzerInnen den Inhalt einer Website bewerten können (Thielsch and Hirschfeld 2019). Insgesamt 12 Fragen decken dabei die vier Bereiche Verständlichkeit, Gefallen, Informationsgehalt und Glaubwürdigkeit ab (die Abkürzung Web-CLIC steht für Website-Clarity, Likeability, Informativeness, Credibility). Zudem kann ein Gesamtwert berechnet werden, dieser spiegelt das subjektive Erleben des Website-Inhalts insgesamt wider. Beispielhafte Items des Web-CLICs sind:

- Die Inhalte sind anschaulich aufbereitet. (Skala: Verständlichkeit)
- Ich lese diese Website gerne. (Skala: Gefallen)
- Die Website ist informativ. (Skala: Informationsgehalt)
- Ich kann den Informationen auf der Website vertrauen. (Skala: Glaubwürdigkeit)

## Referenzen

- Kauermann, G., and H. Küchenhoff. 2011. *Stichproben*. Berlin: Springer.
- Nielsen, J. 1993. *Usability Engineering*. Boston: AP Professional.
- Nunnally, Jum C, Ira H Bernstein, and Jos MF ten Berge. 1967. *Psychometric Theory*. Vol. 226. New York: McGraw-Hill.
- Thielsch, M. T., and G. Hirschfeld. 2019. "Facets of Website Content." *Human Computer Interaction* 34: 279–327. doi:10.1080/07370024.2017.1421954.
- Thielsch, M. T., I. Blotenberg, and R. Jaron. 2014. "User Evaluation of Websites: From First Impression to Recommendation." *Interacting with Computers* 26 (1): 89–112. <http://dx.doi.org/10.1093/iwc/033>.
- Trafimow, D. 2014. "Estimating True Standard Deviations." *Frontiers in Psychology* 5 (235).

# Der optimale Stichprobenumfang für Webevaluationsstudien ist nicht fünf...

...sondern hängt von der gewünschten *Präzision* und der *Reliabilität* des Messinstruments ab.

## Ergebnisse

### Reliabilität und Stichprobenumfang

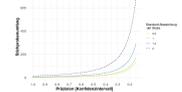


Figure 1: Erforderliche minimale Stichprobenumfänge in Abhängigkeit zur gewünschten Präzision

Abbildung 1 zeigt die benötigten Probanden, um Skalen mit Standardabweichungen von 0,9; 1; 1,3 und 2 unterschiedlich genau zu schätzen. Eine Präzision von 1 bedeutet dabei, dass der wahre Wert nicht weiter als einen Skalenpunkt vom Mittelwert der Ratings entfernt liegt. Ist der Mittelwert der Ratings also bei 4, kann man nur schließen, dass der wahre Wert zwischen 3 und 5 liegt. Das ist für die meisten praktischen Anwendungen wahrscheinlich wenig informativ. Oft ist es daher sinnvoll, eine Präzision von weniger als 0,15 Punkten anzustreben, um den wahren Wert deutlich genauer schätzen zu können – im obigen Beispiel läge der wahre Wert dann zwischen 3,85 und 4,15. Um dies beispielsweise für den Web-CLIC, der eine Standardabweichung von 1,3 aufweist, zu erreichen, benötigt man 290 Probanden.

### Fragebogenlänge und Stichprobenumfang

Im zweiten Schritt verwenden wir nun die Spearman-Brown Formel, um die Reliabilität von unterschiedlich langen Skalen zur Erfassung des Webseiteninhalts zu schätzen. Wir vergleichen dafür auf Basis des Web-CLICs vier verschiedene hypothetische Inhaltsfragebögen mit einer Länge von 1, 2, 4 und 12 Items. Legt man wieder wie im Web-CLIC eine Reliabilität von 0,92 für den 12-Item Fragebogen zugrunde, ergeben sich geschätzte Reliabilitäten für die o.g. Kurzversionen von 0,50; 0,66; 0,76, und 0,92. Diese kann man nun verwenden, um die Standardabweichung der Ratings zu berechnen, aus denen man dann die Anzahl der benötigten Probanden ablesen kann.

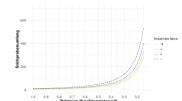


Figure 2: Minimal erforderliche Stichprobenumfänge für unterschiedliche Anzahlen von Items und unterschiedlich hohe Präzision.

Abbildung 2 stellt die erforderlichen Stichprobenumfänge graphisch dar. Wie man sieht, unterscheiden sich die erforderlichen Stichprobenumfänge vor allem dann, wenn die gewünschte Präzision relativ hoch ist. Um ein Konfidenzintervall mit einer Breite von einem Skalenpunkt (= Präzision von 0,5) zu erreichen, sind je nach Anzahl der Items entweder 27 Probanden bei 12 Items oder 49 Probanden bei nur einem Item erforderlich.